



概率统计笔记

奇峰

之前

目录

第一章 随机事件和概率	1
I. 随机事件、古典与几何概型	1
II. 事件关系与概率性质及公式	1
III. 条件概率与乘法公式	3
IV. 独立性与伯努利概型	3
V. 全概率公式与贝叶斯公式	4
第二章 随机变量及其分布	5
I. 随机变量及其分布函数	5
II. 常见分布	6
i. 离散型	6
ii. 连续型	7
III. 随机变量函数的分布	8
i. 离散型随机变量函数的分布	8
ii. 连续型随机变量函数的分布	8
第三章 多维随机变量及其分布	9
I. 二维随机变量及其分布	9
II. 二维离散型随机变量	9
III. 二维连续型随机变量	10
i. 定义与性质	10
ii. 常见分布	11
IV. 二维随机变量函数的分布	12
i. 分布的独立可加性	12
ii. 二维离散型随机变量函数的分布	12
iii. 二维连续型随机变量函数的分布	12
iv. 一离散型一连续型随机变量函数的分布	13
v. 最值函数	13
第四章 随机变量的数字特征	15
I. 数学期望	15
i. 离散型随机变量的数学期望	15



ii.	连续型随机变量的数学期望	15
iii.	数学期望的性质	16
II.	方差	16
III.	随机变量的协方差和相关系数	17
i.	协方差	17
ii.	相关系数	17
iii.	切比雪夫不等式	18
第五章 大数定律和中心极限定理		19
I.	依概率收敛	19
II.	大数定律	19
i.	切比雪夫大数定律	19
ii.	辛钦大数定律	19
iii.	伯努利大数定律	20
III.	中心极限定理	20
i.	列维——林德伯格中心极限定理	20
ii.	隶莫佛——拉普拉斯中心极限定理	20

第一章

随机事件和概率

I. 随机事件、古典与几何概型

随机事件和样本空间

- 样本空间 Ω - 随机试验所有可能结果组成的集合；
- 样本点 ω - 样本空间的元素；
- 随机事件 - 样本空间 Ω 的子集；
- 事件发生 - 当且仅当一子集中一样本点出现时称其发生；

古典概型

若随机试验 E

- 只有有限个样本点 (有限性)；
- 每个样本点出现的可能性相等 (等可能性)；

则称 E 为古典型试验。

若事件 A 中含有 k 个样本点，则其概率为 $P(A) = \frac{A \text{ 样本点个数}}{\Omega \text{ 中样本点个数}} = \frac{k}{n}$ 。

若随机试验 E

- 有无限个样本点；
- 每个样本点出现的可能性相等；

则称 E 为几何型试验。

对事件 A ， $P(A) = \frac{L(A)}{L(\Omega)}$ ，其中 L 代表对应事件的几何度量。

II. 事件关系与概率性质及公式

事件运算的性质

进行事件运算时，一般先逆后积再和差；运算还有性质如下。



- 交换律 - $A \cup B = B \cup A; AB = BA$;
- 结合律 - $(A \cup B) \cup C = A \cup (B \cup C); (A \cap B) \cap C = A \cap (B \cap C)$;
- 分配律 - $(A \cap B)C = (AC) \cap (BC); A \cup (BC) = (A \cup B)(A \cup C)$;
- 德摩根律 - $\overline{A \cup B} = \overline{A} \cap \overline{B}; \overline{A \cap B} = \overline{A} \cup \overline{B}$;

概率的定义、性质与公式

定义 1.2.1 概率的公理化定义

设 E 是一随机事件, Ω 是其样本空间, $P(A)$ 是一映射将每一个事件 A 映射到一实数, 若集合函数 $P\{\bullet\}$ 满足

- 非负性 - 对任意事件 A 有 $P(A) \geq 0$;
- 规范性 - 对必然事件 Ω 有 $P(\Omega) = 1$;
- 可列可加性 - $\forall i \neq j, i, j \in N^*, A_i A_j = \emptyset$, 有 $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$;

则称 $P(A)$ 为事件 A 的概率。

概率有以下性质。

- 非负性 - $\forall A \in \Omega, 0 \leq P(A) \leq 1$;
- 规范性 - $P(\emptyset) = 0; P(\Omega) = 1$;
- 有限可加性 - $\forall i \neq j, i, j = 1, 2, \dots, n, A_i A_j = \emptyset$,
有 $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$;

概率有以下公式。

- 求逆公式 - 对任意事件 $A, P(\overline{A}) = 1 - P(A)$, 常用于正难则反;
- 加法公式 - $P(A \cup B) = P(A) + P(B) - P(AB)$;
- 减法公式 - 对任意二事件 A, B 有 $P(A - B) = P(A\overline{B}) = P(A) - P(AB)$; 特别地, 若有 $B \subset A$, 则有 $P(A - B) = P(A) - P(B)$, $P(B) \leq P(A)$;

概率不等式

- $0 \leq P(A) \leq 1$;
- $B \subset A \Rightarrow P(B) \leq P(A)$;
- $P(A \cup B) \leq P(A) + P(B)$.



III. 条件概率与乘法公式

定义 1.3.1 条件概率

设 A, B 为二事件, 且 $P(A) > 0$, 则称 $P(B|A) = \frac{P(AB)}{P(A)}$ 为在事件 A 发生的条件下, 事件 B 发生的条件概率。

注意, 条件概率满足概率的一切性质。

条件概率具有以下性质。

- $0 \leq P(B|A) \leq 1$;
- $P(\emptyset|A) = 0, P(\Omega|A) = 1$;
- $P(\bar{B}|A) = 1 - P(B|A)$;
- $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A) - P(B_1 B_2|A)$.

计算条件概率时, 抽象问题用定义; 对具体问题, 将概率空间从 Ω 缩小到 A ; 对逆概问题, 利用贝叶斯公式。

条件概率的乘法公式

- 若 $P(A) > 0$, 则 $P(AB) = P(A)P(B|A)$;
- 对事件 A, B, C , 若 $P(AB) > 0$, 则 $P(ABC) = P(A)P(B|A)P(C|AB)$.

IV. 独立性与伯努利概型

独立性

定义 1.4.1 独立性

若 $P(AB) = P(A)P(B)$ 则称事件 A, B 相互独立。

若 A, B 相互独立, 则 \bar{A}, B 和 A, \bar{B} 还有 \bar{A}, \bar{B} 都独立。

三事件独立性

$$A, B, C \text{ 相互独立} \Leftrightarrow \left\{ \begin{array}{l} P(AB) = P(A)P(B) \\ P(AC) = P(A)P(C) \\ P(BC) = P(B)P(C) \\ P(ABC) = P(A)P(B)P(C) \end{array} \right\} \Leftrightarrow A, B, C \text{ 相互独立}$$



若 A, B, C 相互独立, 则 A, B 经过和、积、差运算后得到的事件与 C, \bar{C} 独立, 但是 A, B, C 经过运算的事件不一定。

二事件独立的等价条件

- $P(AB) = P(A)P(B)$;
- $\frac{P(AB)}{P(A)} = P(B) = P(B|A), P(A) > 0$;
- $P(B|\bar{A}) = P(B|A), 0 < P(A) < 1$;
- $P(B|A) + P(\bar{B}|\bar{A}) = 1, 0 < P(A) < 1$;
- $P(B|\bar{A}) + P(\bar{B}|A) = 1, 0 < P(A) < 1$.

n 重伯努利概型

若试验 E 只有 A 和 \bar{A} 两个可能结果, 称 E 为伯努利试验, 每次实验中, $P(A) = p, P(\bar{A}) = 1 - p$; 将伯努利试验独立重复 n 次, 则其中成功 k 次的概率为

$$P_n(k) = C_n^k p^k (1-p)^{n-k}, k = 0, 1, \dots, n$$

V. 全概率公式与贝叶斯公式

完备事件组

若一组事件 A_n 满足

$$\bigcup_{i=1}^n A_i = \Omega, A_i A_j = \emptyset, 1 \leq i \neq j \leq n$$

则称其为完备事件组。

全概率公式

若一组事件 A_n 是完备事件组, 且 $P(A_i) > 0, i = 1, 2, \dots, n$, 则

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

贝叶斯公式

若一组事件 A_n 是完备事件组, 且 $P(B) > 0, P(A_i) > 0, i = 1, 2, \dots, n$, 则

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^n P(A_i)P(B|A_i)}, j = 1, 2, \dots, n$$

第二章

随机变量及其分布

I. 随机变量及其分布函数

将样本空间 Ω 上的实值单值函数 $X = X(\omega), \omega \in \Omega$ 称为随机变量。

$F(x) = P\{X \leq x\}, x \in (-\infty, +\infty)$ 是随机变量的分布函数。

分布函数具有以下性质。

- 非负性 - $0 \leq F(x) \leq 1$;
- 规范性 - $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0; F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$;
- 单调不减性 - $\forall x_1 < x_2, F(x_1) \leq F(x_2)$;
- 右连续性 - $\forall x_0 \in R, F(x_0) = \lim_{x \rightarrow x_0^+} F(x) = F(x_0 + 0)$.

其中规范性可以优先考虑，因为其于微积分有关。

当已知随机变量 X 的分布函数 $F(x)$ 时，有

- $P(X \leq b) = F(b)$;
- $P(X = b) = F(b) - F(b - 0)$;
- $P(X < b) = F(b - 0)$;
- $P(X > b) = 1 - P(X \leq b) = 1 - F(b)$;
- $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$;
- $P(a \leq X < b) = P(X < b) - P(X < a) = F(b - 0) - F(a - 0)$;
- $P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a - 0)$;
- $P(a < X < b) = P(X < b) - P(X \leq a) = F(b - 0) - F(a)$;

其中，前三条的应用最为广泛。

离散型随机变量

离散型随机变量的概率分布形如下表。

其中 $p_k \geq 0, k \in N^*, \sum_{i=1}^{\infty} p_i = 1$.



X	x_1	x_2	\dots	x_k	\dots
P	p_1	p_2	\dots	p_k	\dots

离散型随机变量的分布函数为右连续的阶梯型函数，区间左开右闭，为概率的累加。

连续型随机变量

定义 2.1.1 连续型随机变量概率密度

设随机变量 X 的分布函数为 $F(x)$ ，若存在非负可积函数 $f(x) \geq 0, x \in R$ 使得对任意实数 x ，都有 $F(x) = P\{X \leq x\} = \int_{-\infty}^x f(t)dt$ ，则称 X 为连续型随机变量，函数 $f(x)$ 为 X 的概率密度函数。

定理 2.1.1 $f(x)$ 为密度函数的充要条件

$$f(x) \text{ 是概率密度} \Leftrightarrow \begin{cases} f(x) \geq 0; \\ \int_{-\infty}^{+\infty} f(x)dx = 1. \end{cases}$$

连续型随机变量 X 具有以下性质。

- X 的分布函数是连续函数，因此 $\forall a \in R, P\{X = a\} = 0$;
- $\forall a, b \in R, P\{a < X \leq b\} = \int_a^b f(x)dx$;
- 在 $f(x)$ 的连续点处，有 $F'(x) = f(x)$ 。

对连续型随机变量的题目， $f(x)$ 简单或者具有特殊性质意味着作图解。

II. 常见分布

i. 离散型

离散型随机变量需要注意其取值 (尤其是第一个值) 以及其对应的概率。

0-1 分布

X	0	1
P	$1-p$	p

其中 $0 < p < 1$ 。

二项分布

设事件 A 在任意一次试验中出现的概率均为 $0 < p < 1$ ，而 X 为 n 重伯努利试验中 A 发生的次数，则 X 所有可能取值为 $0, 1, \dots, n$ ，对应的概率为 $P\{X = k\} = C_n^k p^k (1-p)^{n-k}$ 。

几何分布 $G(p)$



若 X 的概率分布为

$$P\{X = k\} = (1-p)^{k-1}p, k = 1, 2, \dots,$$

则称 X 服从参数为 p 的几何分布, 记为 $X \sim G(p)$.

泊松分布 $P(\lambda)$

若随机变量 X 的概率分布满足

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda > 0, k = 0, 1, 2, \dots,$$

则称 X 服从参数为 λ 的泊松分布, 记为 $X \sim P(\lambda)$.

超几何分布 $H(N, M, n)$

若随机变量 X 的概率分布为

$$P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = 0, 1, 2, \min(M, n), M, N, n \in \mathbb{Z}^+$$

则称 X 服从参数为 N, M, n 的超几何分布, 记为 $X \sim H(N, M, n)$.

ii. 连续型

连续性随机变量的密度函数非零区间即其定义区间。

均匀分布 $U(a, b)$

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b, \\ 0, & \text{其他} \end{cases}$$

指数分布 $E(\lambda)$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

注意, 此处可能应用泊松过程的增量平稳性, 即

$$\forall s, t \geq 0, n \geq 0, P\{N(s+t) - N(s) = n\} = P\{N(t) = n\}.$$

正态分布 $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, x \in \mathbb{R}$$

特别地, $X \sim N(0, 1) \Rightarrow \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$, 此时其分布函数为 $\Phi(x)$.

对于一般的正态分布 $X \sim N(\mu, \sigma^2)$, 有 $F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$.

利用正态分布密度函数的规范性, 可求泊松积分 $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$



III. 随机变量函数的分布

i. 离散型随机变量函数的分布

对离散型随机变量函数，采用列表法。

ii. 连续型随机变量函数的分布

对已知概率密度为 $f_X(x)$ 的随机变量 X 有 $Y = g(X)$ ，要求 Y 分布函数 $F_Y(y)$ 和概率密度函数 $f_Y(y)$ 时，有两种办法。

公式法

若 $y = g(x)$ 严格单调，其反函数 $x = h(y)$ 有一阶连续导数，则 $Y = g(X)$ 也是连续型随机变量，其密度函数为

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases} \quad \text{其中 } (\alpha, \beta) \text{ 为 } y = g(x) \text{ 在 } X \text{ 上可能取值的区间上的值域。}$$

分布函数法

先按分布函数的定义求得 Y 的分布函数，再求导得到密度函数。即求 $F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$
 $y) = \int_{g(x) \leq y} f_X(x) dx$ ，再求 $f_Y(y) = F'_Y(y)$ 。

具体而言，连续性随机变量的函数的分布函数法如下。

- i. 由 X 取值范围 (a, b) 确定 y 的取值范围 (c, d) ;
- ii. 由分布函数的定义，确定 F_Y 的左右两头，即对 $F_Y(y) = P(Y \leq y)$,
 - $y < c, F_Y(y) = 0$;
 - $y > d, F_Y(y) = 1$.
- iii. 定中间，即 $y \in (c, d)$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) \\ &\Downarrow \\ &P(X \leq h(y)), \text{此时若 } g(x) \text{ 分段则分段处理} \\ &\Downarrow \\ &\int_{-\infty}^{h(y)} f_X(x) dx, \text{此处取交集} \end{aligned}$$

注意，若连续型随机变量分布函数为 $F(x)$ ，若 $Y = F(X)$ ，则 $Y \sim U(0, 1)$.

第三章

多维随机变量及其分布

I. 二维随机变量及其分布

定义 3.1.1 二维随机变量

设 $X = X(\omega), Y = Y(\omega)$ 是定义在样本空间 Ω 上的两实值单值函数, 则称向量 (X, Y) 为二维随机变量或随机向量。

二维随机变量的分布函数定义为 $F(x, y) = P(X \leq x, Y \leq y)$, 其具有以下性质。

- i. 单调性 - $F(x, y)$ 是变量 x 或变量 y 的单调不减函数, 即对 $x_1 < x_2, y_1 < y_2$, 有 $F(x_1, y) \leq F(x_2, y); F(x, y_1) \leq F(x, y_2)$;
- ii. 有界性 - 对任意 x, y 有 $0 \leq F(x, y) \leq 1$, 且 $F(-\infty, y) = 0; F(x, -\infty) = 0; F(-\infty, -\infty) = 0; F(+\infty, +\infty) = 1$;
- iii. 右连续性 - $F(x, y)$ 分别对 x, y 右连续, 即 $F(x+0, y) = F(x, y) = F(x, y+0)$;
- iv. 非负性 - $\forall x_1 < x_2, y_1 < y_2$, 有 $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$, 即 (X, Y) 落入矩形 $(x_1, x_2] \times (y_1, y_2]$ 区域内的概率。

其具有边缘分布函数, 即

- $F_x(x) = P(X \leq x) = \lim_{y \rightarrow +\infty} F(x, y)$;
- $F_Y(y) = P(Y \leq y) = \lim_{x \rightarrow +\infty} F(x, y)$;

当 $F(x, y) = F_X(x)F_Y(y)$ 时, X, Y 独立。

II. 二维离散型随机变量

二维离散型随机变量的分布为 $P\{X = x_i, Y = y_j\} = p_{ij}, i, j \in N^*$ 。

二维离散型随机变量的题目需要列表, 其事件的概率从表中找对应点。

其边缘密度分布的计算方法为将表按行或列求和, 即

- $p_i = P\{X = x_i\} = \sum_j p_{ij}$;



$$\bullet p_j = P\{Y = y_j\} = \sum_i p_{ij}.$$

X, Y 独立的定义为 $P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}$ 对任意 p_{ij} 成立。

当 $P\{Y = y_j\} > 0$ 时, 在 $Y = y_j$ 的条件下, X 的条件概率为

$$P\{X = x_i | Y = y_j\} = \frac{P\{X = x_i, Y = y_j\}}{P\{Y = y_j\}} = \frac{p_{ij}}{p_j}.$$

Y 的条件概率同理。

注意, 在联合分布列中, 对两随机变量,

- $\exists p_{ij} = 0 \Rightarrow$ 不独立;
- 独立 \Leftrightarrow 分布列按行或列成比例。

III. 二维连续型随机变量

i. 定义与性质

定义 3.3.1 二维连续型随机变量的概率密度

设有二维随机变量 (X, Y) , 其分布函数为 $F(x, y)$, 若存在非负可积的二元函数 $f(x, y)$ 使得对任意实数 x, y 都有 $F(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(u, v) du dv$, 则称 (X, Y) 为二维连续型随机变量, 称 $f(x, y)$ 为其概率密度函数, $F(x, y)$ 是其分布函数。

$f(x, y)$ 具有以下性质。

- 非负性 - $f(x, y) \geq 0$;
- 规范性 - $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$;
- 若 $f(x, y)$ 在 (x, y) 处连续, 则有 $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$.

随机点落在区域 G 内的概率为 $P\{(X, Y) \in G\} = \iint_G f(x, y) dx dy$

X 的边缘概率密度为 $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$, 对 Y 同理。

若 X, Y 独立, 则 $f(x, y) = f_X(x)f_Y(y) \forall x, y$.

对给定的实数 y , Y 的边缘概率密度 $f_Y(y) > 0$, 则此时 X 的条件概率密度为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

同理可以得到 Y 的条件概率密度。

注意, 对 x, y 使得 $f_Y(y) > 0$, 有 $f_{X|Y}(x|y) = \frac{f(x)f_Y(y)}{f_Y(y)}$.



对二维连续随机变量区域划分时，需要划分的是概率密度非零区域。需要找到边界点，然后向上向右作射线。

对二维连续随机变量，

$$X, Y \text{ 独立} \Leftrightarrow \begin{cases} f(x, y) \text{ 非零区域为矩形区域;} \\ f(x, y) \text{ 变量可分离.} \end{cases}$$

ii. 常见分布

二维均匀分布

若 (X, Y) 的概率密度为 $f(x, y) = \begin{cases} \frac{1}{S_D}, & (x, y) \in D \\ 0, & (x, y) \notin D \end{cases}$ ，则称 (X, Y) 服从区域 D 上的二维均匀分布。

注意，

- i. 对区域 $G \subset D$ ，有 $F = \frac{S_G}{S_D}$;
- ii. 二维均匀分布的边缘分布为一维均匀分布，当且仅当其非零区域为矩形；
- iii. 二维均匀分布的条件分布一定是均匀分布。

二维正态分布

若 (X, Y) 概率密度为

$$f(x, y) = \exp \left\{ -\frac{1}{2(1-\rho^2)} \right\} \left\{ \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} \right\} / \left\{ 2\pi\sigma_1\sigma_2\sqrt{1-\rho^2} \right\}$$

其中 $x, y \in R, \mu_i, \sigma_i, \rho_i > 0, -1 < \rho < 1$ 均为常数，则称 (X, Y) 服从参数为 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 的二维正态分布，记为 $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$ 。

二维正态分布具有性质如下。

- X, Y 独立 $\Rightarrow \rho = 0$;
- 两个边缘分布服从一维正态分布，即 $X \sim N(\mu_1, \sigma_1^2)$ 和 $Y \sim N(\mu_2, \sigma_2^2)$ ，且 ρ 为二者相关系数；
- X, Y 的任意非零线性组合 $aX + bY$ 服从一维正态分布，即

$$aX + bY \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\sigma_1\sigma_2\rho)$$

- 若 $Z_1 = aX + bY$ 与 $Z_2 = cX + dY$ 为 X, Y 的非零线性组合，若 $\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0$ ，则 (Z_1, Z_2) 仍然服从二维正态分布。



IV. 二维随机变量函数的分布

i. 分布的独立可加性

- 若 $X \sim B(m, p), Y \sim B(n, p)$, 且 X, Y 相互独立, 则 $X + Y \sim B(m + n, p)$;
- 若 $X \sim P(\lambda_1), Y \sim P(\lambda_2)$, 且 X, Y 相互独立, 则 $X + Y \sim P(\lambda_1 + \lambda_2)$;
- 若 $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, 且 X, Y 相互独立, 则 $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$;

更一般地, 若 $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \dots, n$, 且 $X_i, i = 1, 2, \dots, n$ 相互独立, 则

$$Y = \sum_{i=1}^n C_i X_i + C \sim N\left(\sum_{i=1}^n C_i \mu_i + C, \sum_{i=1}^n C_i^2 \sigma_i^2\right)$$

其中 $C_i, i = 1, 2, \dots, n$ 是不全为零的常数。

ii. 二维离散型随机变量函数的分布

使用表格法列出关键取值对及其对应的概率。

iii. 二维连续型随机变量函数的分布

已知二维连续型随机变量 (X, Y) 的概率密度 $f(x, y)$, 求连续函数 $Z = g(X, Y)$ 的概率密度 $f_Z(z)$.

可以采取分布函数法或公式法。

线性规划最值时, 注意可行域的边界点。

分布函数法

对于分布函数法, 具体而言,

- 由 $(x, y) \in D$ 得到 $z = g(x, y) \in [a, b]$;
- 由 $F_Z(z) = P(Z \leq z), z \in R$ 定两边, 即 $z < a, F_Z(z) = 0; z > b, F_Z(z) = 1$;
- 对 $z \in [a, b]$,

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = P(g(X, Y) \leq z) \\ &= \iint_{g(x, y) \leq z \cup D} f(x, y) dx dy \end{aligned}$$

注意, 不要忘记与非零区间取交集。



公式法

对于公式法, 设 (X, Y) 的联合概率密度为 $f(x, y)$, 有以下公式。

求和 $Z = X + Y$

$Z = X + Y$ 的概率密度为

$$f_z(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx = \int_{-\infty}^{+\infty} f(z-y, y)dy$$

若 X, Y 还相互独立, 则适用卷积公式

$$f_z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx = \int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)dy$$

求差 $Z = X - Y$

$Z = X - Y$ 的概率密度为

$$f_z(z) = \int_{-\infty}^{+\infty} f(x, x-z)dx = \int_{-\infty}^{+\infty} f(z+y, y)dy$$

求积 $Z = XY$

$Z = XY$ 的概率密度为

$$f_z(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x})dx$$

求商 $Z = \frac{X}{Y}$

$Z = \frac{X}{Y}$ 的概率密度为

$$f_z(z) = \int_{-\infty}^{+\infty} |y| f(yz, y)dy$$

公式法的要点

以求和为例, 对 $f(z) = \int_{-\infty}^{+\infty} f(x, z-x)dx$, z 变动意味着非零区间变动, 此时非零区间为 $(x, z-x) \in D$.

iv. 一离散型一连续型随机变量函数的分布

结合全概率公式对离散型变量进行全集分解, 也即分类讨论。

v. 最值函数

对极大值 $U = \max(X, Y)$, 当 X, Y 独立同分布时, $F_U(u) = [F_x(u)]^2$, 因此密度为 $f_U(u)F'_U(u) = 2F_X(u)f_X(u)$.

对极小值 $U = \min(X, Y)$, 当 X, Y 独立同分布时, $F_U(u) = 1 - [1 - F_x(u)]^2$, 因此密度为 $f_U(u)F'_U(u) = 2[1 - F_X(u)]f_X(u)$.



事实上，这是顺序统计量：对于一组 n 个独立同分布的随机变量，这组中第 k 大的密度函数为

$$f_k(x) = n! \frac{[F(x)]^{n-1}}{(n-1)!} \frac{[1-F(x)]^{n-k}}{(n-k)!} f(x)$$

据此可求分布函数。

注意，若 $X \sim E(\lambda_1), Y \sim E(\lambda_2)$ 且二随机变量独立，又有 $Z = \min(X, Y)$ ，则 $Z \sim E(\lambda_1 + \lambda_2)$ 。

对二维随机变量 (X, Y) ，令 $U = \max(X, Y), V = \min(X, Y)$ ，则有

- $U + V = X + Y$;
- $U - V = |X - Y|$;
- $UV = XY$.

第四章

随机变量的数字特征

I. 数学期望

i. 离散型随机变量的数学期望

一维离散型随机变量的数学期望

$$E(X) = \sum_{i=1}^{\infty} x_i p_i$$

一维离散型随机变量函数的数学期望

$$E(g(X)) = \sum_{i=1}^{\infty} g(x_i) p_i$$

二维离散型随机变量的数学期望

$$E(g(X, Y)) = \sum_{i=1}^{\infty} \sum_{j=1}^{+\infty} g(x_i, y_j) p_{ij}$$

ii. 连续型随机变量的数学期望

一维连续型随机变量的数学期望

$$E(x) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} x dF(x)$$

一维连续型随机变量函数的数学期望

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x) dx = \int_{-\infty}^{+\infty} g(x) dF(x)$$

二维连续型随机变量的数学期望

$$E(g(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy$$



iii. 数学期望的性质

对常数 c, c_1, c_2 , 随机变量 X, Y ,

- i. $E(c) = c$;
- ii. $E(cX) = cE(X)$;
- iii. $E(c_1X + c_2Y) = c_1E(X) + c_2E(Y)$;
- iv. 若 X, Y 独立, 则 $E(XY) = E(X)E(Y)$.

II. 方差

定义 4.2.1 方差

$$D(x) = E[X - E(X)]^2.$$

计算方差时, 常用的公式为

$$D(x) = E(X^2) - [E(X)]^2$$

对常数 a, b , 随机变量 X, Y , 有

- $D(c) = 0$;
- $D(aX + b) = a^2D(X)$;
- $D(X \pm Y) = D(X) + D(Y) \pm 2\text{Cov}(X, Y)$, 其中 $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

常见分布的数学期望和方差如下表。

分布名称	分布记号	数学期望	方差
0-1 分布	$X \sim B(1, p)$	p	$p(1 - p)$
二项分布	$X \sim B(n, p)$	np	$np(1 - p)$
泊松分布	$X \sim P(\lambda)$	λ	λ
几何分布	$X \sim G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
超几何分布	$X \sim H(n, M, N)$	$n \frac{M}{N}$	
均匀分布	$X \sim U(a, b)$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
指数分布	$X \sim E(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布	$X \sim N(\mu, \sigma^2)$	μ	σ^2
卡方分布	$\chi^2 \sim \chi^2(n)$	n	$2n$

注意事项

当给定一个含参数的概率密度函数时, 向常见分布上凑。



对求 $E(g(x))$ ，可以将其化为 $\int_{-\infty}^{+\infty} g(x)f(x)dx$ ，然后尝试将其整理为 $\int_{-\infty}^{+\infty} g_1(x)f_1(x)dx$ ，其中 $f_1(x)$ 是一常见分布 T 的密度函数，从而将其化为 $E(g_1(T))$ 。

对于求分布函数 $F(x)$ 为不同分布的分布函数和的随机变量的期望，将期望转化为 $\int x dF(x)$ ，将 $F(x)$ 拆开，分别计算积分。

有时，可以利用等式 $E(X^2) = E(X(X-1) + X)$ 。

III. 随机变量的协方差和相关系数

i. 协方差

定义 4.3.1 协方差

$$\text{Cov}(X, Y) = E[X - E(X)][Y - E(Y)].$$

计算协方差时，常用公式 $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ 。

协方差有如下性质。

- $\text{Cov}(X, X) = D(X)$ ；
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ ；
- $\text{Cov}(aX + b, cY + d) = ac \cdot \text{Cov}(X, Y)$ 。

ii. 相关系数

定义 4.3.2 相关系数

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}.$$

$\rho_{XY} = 0$ 时称 X, Y 不相关，否则称其相关。

注意，独立 \Rightarrow 不相关，特别地，对二维正态分布的两边缘分布有不相关性 \Rightarrow 独立性。

对随机变量 X, Y ，相关系数有如下性质。

- $|\rho_{XY}| \leq 1$ ；
- $\exists a \neq 0, b, P\{Y = aX + b\} = 1 \Leftrightarrow \rho_{XY} = \frac{|a|}{a}$ 。

注意事项

对于难以快速计算的题目，应用验证法，即通过已知性质排除备选项。



iii. 切比雪夫不等式

定理 4.3.1 切比雪夫不等式

设随机变量 X 期望和方差 $E(X) = \mu, D(X) = \sigma^2$ 都存在, 则对任意 $\varepsilon > 0$ 都有

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}$$

证明

$$\begin{aligned} P\{|X - E(X)| \geq \varepsilon\} &= \int_{|X - E(X)| \geq \varepsilon} f(x) dx \\ &\leq \int_{|X - E(X)| \geq \varepsilon} \frac{|X - E(X)|^2}{\varepsilon^2} f(x) dx \quad (\text{放大被积函数}) \\ &\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (X - E(X))^2 f(x) dx \quad (\text{放大积分区间}) \\ &= \frac{1}{\varepsilon^2} E((X - E(X))^2) = \frac{D(X)}{\varepsilon^2} \end{aligned}$$

整理即为待证结论。 ■

第五章

大数定律和中心极限定理

I. 依概率收敛

定义 5.1.1 依概率收敛

设 $X_i, i = 1, 2, \dots, n$ 是一列随机变量序列, a 是一实数, 若对任意正数 ε 都有 $\lim_{n \rightarrow +\infty} P\{|X_n - a| \leq \varepsilon\} = 1$, 则称上述序列依概率收敛于 a , 记为 $X_n \xrightarrow{P} a$.

依概率收敛有如下性质。

- 若 $X_n \xrightarrow{P} X$, 而 $g(x)$ 是 R 上的连续函数, 则 $g(X_n) \xrightarrow{P} g(X)$.
- 若 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$ 而 $g(x, y)$ 在 (a, b) 处连续, 则 $g(X_n, Y_n) \xrightarrow{P} g(a, b)$.

II. 大数定律

大数定律的本质是算数平均依概率收敛于统计平均, 也即数学期望。

i. 切比雪夫大数定律

设 $X_i, i = 1, 2, \dots$ 其中任意的两随机变量独立, 若其方差 DX_i 均存在且有一致的上界, 也即存在 $C \in R$ 使得 $DX_i \leq C$, 则

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i\right| \leq \varepsilon\right\} = 1$$

事实上, 有 $\frac{1}{n} \sum X_i \xrightarrow{P} \frac{1}{n} \sum EX_i$.

ii. 辛钦大数定律

设随机变量 $X_i, i = 1, 2, \dots$ 相互独立同分布, 且其数学期望 $EX = \mu$, 则对任意 $\varepsilon > 0$ 都有

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \leq \varepsilon\right\} = 1$$



iii. 伯努利大数定律

设 n_A 是 n 次独立重复试验中 A 发生的次数, p 是事件 A 在每次试验中发生的概率, 则

$$\lim_{n \rightarrow +\infty} P \left\{ \left| \frac{n_A}{n} - p \right| \leq \varepsilon \right\} = 1$$

事实上, 令 $X_i = \begin{cases} 1, & A \\ 0, & \bar{A} \end{cases}$, 其中 $X_i \stackrel{i.i.d}{=} X_j, i \neq j$, 容易发现其是辛钦大数定律的推论。

III. 中心极限定理

其本质为足够多的 ($n \geq 30$) 相互独立的随机变量和近似地服从正态分布。

i. 列维—林德伯格中心极限定理

设随机变量 $X_i, i = 1, 2, \dots$ 相互独立同分布, 有 $EX = \mu, DX = \sigma^2$, 则当 $n \rightarrow \infty$ 时有

$$\begin{aligned} \sum_{i=1}^n X_i &\sim N(n\mu, n\sigma^2) \\ \Rightarrow \frac{\sum X_i - \sum EX_i}{\sqrt{\sum DX_i}} &\sim N(0, 1), n \rightarrow \infty \\ \Rightarrow \lim_{n \rightarrow \infty} P \left\{ \frac{\sum X_i - \sum EX_i}{\sqrt{\sum DX_i}} \leq x \right\} &= \Phi(x) \end{aligned}$$

ii. 隶莫佛—拉普拉斯中心极限定理

设随机变量 X_n 服从参数为 n, p 的二项分布, 则当 $n \rightarrow \infty (\geq 30)$ 时, 有

$$X_n \sim N(np, np(1-p))$$